DOCUMENT RESUME

ED 221 581 TM 820 611

AUTHOR Winters, Lynn
TITLE The Role of Evaluation and Plans for Evaluating the Current Testing Program.
INSTITUTION Palos Verdes Peninsula Unified School District, CA.
PUB DATE 21 Sep 81
NOTE 33p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary Secondary Education; Evaluation Criteria; *Evaluation Methods; Evaluators; Minimum Competency Testing; *Program Effectiveness; Program Evaluation; *Research and Instruction Units; Standardized Tests; Test Construction; *Testing Programs; Test Validity
IDENTIFIERS *Palos Verdes Peninsula Unified School District CA

ABSTRACT

The Palos Verdes Peninsula Unified School District Office of Program Evaluation and Research is responsible for providing information for program development and improvement; providing test information to special programs coordinators; and acting as a clearinghouse for all information concerning tests, evaluation methodology, and educational research. This report defines the evaluation process used; and describes the activities of evaluators, the types of evaluations conducted, and the role of the office of evaluation and research. The standardized testing program is evaluated to determine its adequacy and potential improvements in the efficient use of test results. The relevance of the testing program to the curriculum is also evaluated. The competency testing program was reviewed to compare costs for development of an in-house test, purchase of a published test, commission of Riverside Publications to produce a customized test, and adaptation of a testing program developed by another district. The general plan for evaluating the testing program in terms of adequacy, relevance, and cost-effectiveness is presented. (DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED221581

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
L. Winters
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Palos Verdes Peninsula Unified School District

THE ROLE OF EVALUATION AND PLANS FOR EVALUATING THE CURRENT TESTING PROGRAM

by

Lynn Winters, Director of Program Evaluation and Research

Part 1

WHAT IS EVALUATION AND HOW SHOULD IT AFFECT THE INSTRUCTIONAL PROGRAM?

This report has two purposes: The first is to describe what evaluation is and how it affects the instructional program. The second is to present the evaluation and research office's plan for evaluating the current testing program in terms of adequacy, relevance, and cost-effectiveness.

The Function of the Office of Program Evaluation and Research

The Office of Program Evaluation and Research is responsible for providing administrative and school staff with information for program development and improvement. It also provides test information to coordinators of special programs, such as School Improvement, Bilingual, and Special Education, so that they can complete reports required by state and federal funding agencies.

Finally, it is the clearinghouse for information about tests, evaluation methodology, and educational research. In order to understand how this office fulfills these three responsibilities, it is important to know what evaluation is. This section will answer four basic questions that define evaluation:

1. What is evaluation?
2. What do evaluators do?
3. What kinds of evaluations are conducted in this district?
4. What is the role of the office of evaluation and research?

TM 820 611

What is Evaluation?

Evaluation is the collection of information for decision-making. The evaluator asks the decision-maker two questions that determine what information is to be gathered during the evaluation.

What do you need to know that you do not already know?

What will this information allow you to do that you cannot do now?

The answers to these two questions define the purposes of an evaluation. After the focus is established, the evaluation is conducted according to the familiar "scientific method." The evaluator

- identifies the questions/hypotheses guiding the evaluation
- identifies the sample from whom information will be gathered
- selects/develops instruments to get desired information
- uses the instruments (surveys, observations, tests, case studies) to collect data
- summarizes the data both analytically and statistically
- presents results in a form easily understood by decision-makers

Evaluation, although it incorporates much of the same methodology as research, is not basic research. It does not always provide results that will generalize to other students and programs. Evaluation, although it often uses tests, is not testing. Good evaluations rely on many different sources of information for making conclusions. Evaluation, although it uses standardized data collection procedures, is not completely objective. Its goals and findings emerge directly from a district's own priorities and values. Evaluation, although it provides empirical information, is not appropriate for all decisions. If the information gathered in an evaluation cannot be used to make <u>better</u> decisions, then evaluation is an aimless and irrelevant activity.

Many kinds of evaluation. Not all evaluations have the same goals. Depending on the kind of program decision that must be made, an evaluation may answer any of five questions. Table 1 shows the different kinds of evaluation activities associated with the five major program decisions faced by a school district.

TABLE 1.

Evaluation Activities Associated with Different Kinds of Program Decisions

Decision	Evaluation Question	Evaluation Activity
Student needs	What are the strengths and weaknesses of district students?	Needs Assessment
Program development	What kinds of programs meet student needs?	Program Planning
Program Adoption	Are the programs being adopted as planned?	Implementation Evaluation
Program Improvement	What is working well/ poorly in the program?	Formative Evaluation
Program Worth	Is the program good?	Summative Evaluation

Although the general methods of conducting evaluations are the same, the kind of information collected for each of the questions above is different. Some evaluations, such as needs assessment or summative evaluation, are guided by prespecified decisions about what students should know and how well they should be performing for the program to be considered "good." Others, such as program planning and formative evaluation, are intended to describe programs so that decisions about district priorities and program outcomes can be made.

What Do Evaluators Do?

When conducting an evaluation of any type, the evaluator plays several roles. In the early stages of a study, the evaluator acts as a facilitator among those interested in the evaluation results, coaxing from them the goals and political consequences of the evaluation. Next, the evaluator designs an information collection plan reflecting valid research design. The evaluator then selects or develops instruments to get desired information.

Once the instruments have been designed, observations of the sample are made. Observations may consist of administering standardized achievement tests, collecting samples of student work, visiting classrooms, interviewing teachers, or surveying school staff and parents. The information thus gathered is organized, either manually or by computer, then summarized using logical or statistical analyses.

Finally, the evaluator reports the findings to the people responsible for making decisions about students and programs. Because of the many different kinds of skills involved in conducting a successful evaluation, good evaluators involve several professionals, teachers, administrators, testing specialists, and data processing personnel, in the planning and implementation of an evaluation.

What Kinds of Evaluations are Conducted in PVPUSD?

Palos Verdes gathers evaluation information for two general audiences: external groups such as state and federal funding agencies, and internal groups such as school staff, district administration, and the Board of Education. Table 2 A-B shows which programs report to external or internal audiences, and the kinds of information these programs are required to gather. Although several evaluations of programs will be conducted in 1981-82, data collected for one may often be used to fulfill another's reporting requirements as well.

TABLE 2. A

The Evaluation Reporting Requirements of District Programs: 1981-82

EXTERNAL REPORTS
(State and Federal Funding Agencies)

Program	Test Data Needed
Bilingual	Local norms for the non-minority population on standardized tests of reading, English language, and mathematics, grades K-12
	Average normal curve equivalent score for all LEP and FEP students in the areas of reading, language arts, and mathematics
	Individual student scores for each LEP student on standardized reading, language, and mathematics tests
Title I	Percentile score on standardized reading test used for screening Normal curve equivalent score (pre and post) on standardized test used for testing reading (The screening test score cannot be used for the pretest).
Special Education	Competency test scores for all students whose IEP's require competency testing Percentile ranks in reading, language arts, mathematics on a standardized screening test Measures of attainment of course objectives as specified in IEP

TABLE 2 B

INTERNAL REPORTS

(School Staff, District Administration, School Board, Parents)

Program	Test Data Needed
Standardized Testing Program	Individual student scores on each section of the Iowa Tests of Basic Skills (grades K-8) or the Iowa Tests of Achievement and Proficiency (grades 9-12)
	Norm-referenced comparisons for individual students, schools, and the district for ITBS/I
	California Assessment Test: percent correct scores for schools and district (grades 1, 3, 6, 12)
	Percentile ranks for California school districts on the CAP
	School performance on the CAP within the "comparison score band"
Competency Testing Program	Individual students' patterns of pass/fail on the 16 district competencies
	Item analyses for each of the 16 tests at the student, classroom, and school level
	Per cent pass rate on each test, and for all tests combined, for grades 5, 8, and 9-12
	Names of students not graduating because of competencies
Writing Standards Program	Scores on the writing sections of the competency test for grades 5, 8, 9-12
	Evaluation of 11th grade writing for each student for composition skills placement
	Evaluation of selected samples of writing at each grade level to assess impact of new program
Gifted and Talented Program	Percentile ranks on each section of a standardized achievement battery for screening (gr.
	Comparison of GATE performance on ITBS, ITAP, and competency with regular student performance
Reading Specialist Program	Percentile rank/scale score on a standardized reading test
	Pre-post scale or normal curve equivalent scores for measuring program impact
	Diagnostic information (may or may not be standardized test)
Bilingual Program	Writing scores (reliably rated) for all students seeking reclassification to FEP
	Information that allows bilingual personnel to decide when a LEP student is capable of "competing successfully" (getting a C) in a regular (not modified) program.

The Office of Program Evaluation and Research is responsible for gathering the information required by three programs: standardized testing, competency, and writing standards. Because test results from these three programs are not available for all students at all grade levels, staff members coordinating other programs, such as Title I or Bilingual, must administer their own tests for evaluation and reporting. Since each program has operated rather independently from the other, there has been a proliferation of standardized tests in this district. In addition to the Iowa test and CAP, teachers administer the Gates McGinitie, Stanford Diagnostic, Metropolitan, and old versions of the Comprehensive Tests of Basic Skills.

What is the Role of the Office of Program Evaluation and Research?

The role of this office is to provide information required by district staff for external and internal evaluation and reporting. This information was not always provided because of the high cost of data processing associated with district testing. The major goal of this office in 1981 is to evaluate the district testing program, consisting of standardized tests, competency tests, and writing standards tests in terms of adequacy, relevance, and cost. Specifically, this office is reviewing the testing program to determine:

How can test results be used more efficiently?

How well does the present program match the curriculum?

How can testing and data processing costs be reduced?

Part II of this report describes how this office is answering these three questions and what concerns it has with regard to each of the three programs in terms of their adequacy, relevance, and cost.

Part II

EVALUATION OF THE CURRENT Testing Program

Adequacy: How Can Test Results be Used More Efficiently?

Table 3 describes the Testing Schedule for 1981-82. The basic question of test adequacy is: will the district testing program supply all information needed to complete internal and external evaluation requirements for 1981-82? A review of the current schedule says, "no." Standardized test results for screening and external reporting will be available only for grades 1-5, 8, and 12. Test score information needed at other grade levels must be obtained by individual programs. Thus, Bilingual, Title I and Special Education must either administer and score the Iowa themselves or select another test to use. If the same test were used by all of these special programs, as well as for the district standardized testing program, test administration, purchase, and scoring costs would be substantially reduced on a per pupil basis. Further, the availability of comparable scores for students in special and regular programs would make between-program comparisons possible and allow the validity of special program placement to be examined.

Coordination of district testing and evaluation activities with the above purposes in mind has begun under the direction of Fred London and the School Improvement Program. Staff representing all programs requiring standardized test information met in early September, 1981, to document program placement criteria and reporting requirements. This information is being gathered by the SI office for the October meeting of special program coordinators. At that meeting two questions will be answered:

1. Can a single standardized test serve all internal and external reporting requirements?
2. If so, which tests should be given, at what grades, and when?

TABLE 3.

Testing Schedule: 1981-82

Test	Purpose(s)	Grades	Date(s)
California Assessment Program	Program evaluation State monitoring of California basic skills program	1	9/28-10/9
		3	4/21-5/11
		6	4/21-5/4
		12	12/2-12/15
Iowa Tests of Basic Skills	State and federal reporting for special programs Screening for GATE, modified, special education Report to parents on student performance compared with a national sample	1,2,3,4,5	3/3-3/21
Iowa Tests of Achievement and Proficiency	Same purposes as ITBS	12	12/2-12/15
Competency Tests Reading and Mathematics	Assess student progress toward district minimal competencies	5	5/3-23
	Certification of reading and math competence for graduation	8	4/22-23
		10	3/3
		11	3/2
		12	12/9, 3/2, 3/27
Writing Standards Appeals	Allow students failing the 5/31 essay test for placement in grade 12 composition skills to retake the test	11-12	9/23
Writing Standards (Same as Writing Competency Test)	Assess student progress on district writing competencies	5	4/21-22
	Certification of minimal competence in writing	8	4/26
		10-12	Same as Competency above

Standardized

Competency

Writing Standards

Following the October meeting, the Office of Program Evaluation and Research will guide selection of the appropriate test(s) and work with program coordinators to draft the 1982-83 testing schedule.

Relevance: How Well Does the Present Testing Program Match the Curriculum?

Table 4 describes the two kinds of tests administered in PVPUSD: norm-referenced and criterion-referenced. A norm-referenced test is one that reports how well a student is doing compared to other students who took the same test. A criterion-referenced test reports the extent to which students have mastered instructional objectives. The standardized testing program uses norm-referenced tests. The competency testing and writing standards program require criterion-referenced tests.

Norm-referenced tests. Norm-referenced tests are constructed to rank students in comparison to their grade or age-level peers. Test items are written to "spread" students out into an approximately normal (bell-shaped) curve so that percentile ranks may be calculated. Any test item that all students answer correctly or that all students miss is deleted from the test *regardless of its usefulness in measuring curricular objectives.*

On a norm-referenced test, the "average" student is supposed to answer from 45% to 65% of the items correctly. Scores are not valid for students who answer less than 30% or more than 75% of the items. These students are said to have reached the test "floor" or "ceiling" and are considered outside of the functional range of the test. Students who score at the floor should be given an easier level of the test; those who score at the ceiling, a harder level.

TABLE 4.

A Description of the District Program: Tests and Purposes

Norm-Referenced Tests

California Assessment Tests

Purposes: Report to California Legislature by State Department of Education
Allow districts to assess progress on state basic skills objectives
Allow districts to compare themselves with "like" districts

Content: Grade 1: Reading readiness
3: Reading, written expression, mathematics
6: Reading, written expression, mathematics
12: Reading, written expression, mathematics

Time: 35-45 minutes

Scores: Percent correct for each area and subskills within areas for the district and each school within a district
State percentile ranks for schools and district
District percentile comparison with "like" districts

Iowa Tests of Basic Skills and Iowa Tests of Achievement and Proficiency

Purposes: Competency assessment (ITBS, grades 5 and 8)
Screening for special programs (grades 1-3)
State and federal reporting (grades 1-12)
Reports to parents

Content: Grades K-8 (ITBS): Vocabulary
Reading Comprehension
Language Skills
Work Study Skills
Mathematics Skills

Grades 9-12 (ITAP) Reading Comprehension
Mathematics
Using Sources of Information
Written Expression

Time: Two-three days (ITBS)
Four hours (ITAP)

Scores: Number or percent correct for each subskill and test
National individual, school, and district percentile ranks
Special norms: High Achieving Schools; Low Achieving Schools

TABLE 4. (cont.)

Criterion-Referenced Tests

Competency Tests

Purposes: Early detection of students not proficient in minimal competencies
Certification for graduation from high school

Content: Grade 5: Reading and math tests--subset of the Iowa tests, level 9
Writing Test--same as grade 5 writing standards

- Filling out a form
- Note taking
- Paragraph (6-10 sentences)
- Personal Letter

Grade 8: Reading and math tests--subset of Iowa tests, level 14
Writing Test--same as grade 8 writing standards

- All of the above writing skills plus,
- Bibliography
- Business letter
- Outlining
- Sentence combining

Grades 9-12: District-constructed reading, writing and math forms

Scores: Percent of items correct on each of the 16 subtests
Pass/fail scores: Pass =50% correct

Writing Standards Tests

Purposes: Monitor student growth in writing
Determine placement in remedial composition (grade 12)

Content: Grades 5 and 8: Same as the writing competency tests
11: Five-paragraph essay

Scores: Number correct assigned according to a standardized scoring guide
Pass/fail scores: grades 5, 8 Pass=50%
grade 11 Pass =55%

Because these tests are used nation-wide, they do not measure the instructional objectives of any one school district. They are not intended for local program evaluation or student diagnosis on instructional goals. Norm-referenced tests have two valid uses: screening, to identify low and high scoring students for possible special program placement; comparison, to measure the extent to which district students are above or below their peers on objectives common to a hypothetical "national" curriculum. Use of the CAP or Iowa tests to measure district objectives is not valid. Use of these scores to compare students or schools with other students or schools is.

Concerns about the standardized testing program. The Office of Program Evaluation and Research has four concerns about the relevance of the standardized, norm-referenced testing program:

1. What, if any, kind of program evaluation information should be obtained from these two tests?
2. Who should get test results, when, and in what format?
3. Which grade levels should be tested and when?
4. Which norms are appropriate for comparing Palos Verdes students to others; national norms, high achieving norms, etc.

This office also has a concern about the scoring programs currently used. In the past, the district has provided teachers with grade equivalent and percentile scores. Grade equivalent score reporting is a dubious practice, and one that is discouraged by the American Psychological Association, the American Educational Research Association, National Council on Measurement in Education, Office of Education and CTB-McGraw Hill, among others. Neither grade equivalent nor percentile ranks are useful for program evaluation or for monitoring the gains of groups of students over time. Useful scores for teachers

tend to be raw scores (number correct), percent correct, and percentile ranks. A second concern about the scoring programs is with the practice of using an average percentile rank for each grade level and with the method used in calculating a percentile rank for the district. In the past, the average of the individual percentile ranks of students at the same grade in a school or the district have been calculated. This practice is questionable. Correct calculation of these scores requires that the individual *scale scores* be averaged and then a separate norms table for schools or districts be consulted to get the school or district percentile rank. These tables are available from Iowa but must be entered into the scoring program.

Criterion-referenced tests. Evaluation of the competency testing program (including writing standards) focuses on different sorts of relevance issues than does review of the standardized testing program. The competency and writing standards tests have criterion-referenced interpretations. They are supposed to match district objectives; their scores indicate how well students have mastered these objectives. Test construction methods for criterion-referenced tests differ considerably from those used for norm-referenced tests. Items are written, not to discriminate among individuals, but to reflect the content and skills related to a particular objective. Thus, item writing is guided by detailed content specifications, and many items are used to measure one objective. While a criterion-referenced test may contain norms, its chief purpose is not to rank students; it is to describe their performance. Norm-referenced tests, in contrast, do not contain enough items per objective to yield criterion referenced interpretations.

Curricular relevance of the competency and writing standards tests.

The use of the Iowa tests for competency assessment in reading and mathematics

for grades 5 and 8 is a suspect practice for several reasons:

1. Test items were not designed to measure local objectives and often assess skills not taught in school.
2. There are not enough items per skill to make reliable estimates of students' competency levels.
3. The same test forms have been used for 2-3 years which means they probably are not secure.
4. Students are asked to answer many irrelevant items in order to "get to" questions that "count" for competency assessment.
5. It is questionable that students do their best when the tests are administered four hours a day for three consecutive days (grade 8).

Additionally, the 9-12 competency test forms need review. These "make-up" forms, that 55-65% of the eighth grade students will be taking sometime in grades 9-12, are a cut-and-paste version of Iowa items from old versions of the ITBS (levels 13-14). The development of a reliable and valid test, however, presupposes much more than pasting together new forms. This process is outlined in Table 5 and takes about two years. Omitting any of these development steps seriously compromises the psychometric quality of the test.

The writing standards test is used for assessing writing competence. As such, it is part of the competency program. Because of its local origin and systematic scoring criteria, it is probably a valid example of a proficiency test. However, the district has no information about test reliability, nor does it have specifications for the writing test items. Clear item specification is crucial for assessing the content validity of a test and for constructing new items. This office plans to study writing test reliability and supply item specifications for the writing standards forms used in competency testing.

TABLE 5.

Steps in the Test Development Process

Task	Amount of Time
1. Development of test specifications	8 months
a. Define purposes of test	
b. Create content specifications	
c. Select item formats	
d. Determine number of items per skill	
e. Estimate cognitive difficulty of items	
2. Item Production and Revision	9 months
a. Write items to fit specifications	
b. Edit and try-out items with students	
c. Revise items and try-out	
d. Construct test forms	
e. Try-out test forms	
f. Revise and try-out forms	
g. Calculate item and test statistics	
3. Preparation of Score Interpretation Information	4 months
a. Conduct norming or standard-setting study	
b. Write technical manual describing test use	
4. Printing and Production of Test	3 months
a. Paste-up, layout	
b. Proofing of galleys	
c. Final printing	

Finally, the question of whether the 16 competencies are being taught must be investigated. Competencies tested in grades 5 and 8 are taught in conjunction with the elementary and intermediate school basic skills program. It is not clear, however, where the competencies fit into the high school curriculum. The district does provide a summer workshop for those students in grades eight and beyond who have failed the competency tests. Nevertheless, a survey conducted in July, 1981, indicated that of 382 students completing the survey (about 450 were tested), over 50% said they had "no special preparation" or "studied on their own." Sixty-eight percent of this group (382) had just taken the test as eighth graders in May. The obvious question: Do students need remediation to pass the tests, and if so, when do they get it?

The Importance of Competency Testing to PVPUSD. A final question related to the competency testing program is one of district commitment. Last year, competency testing cost the district approximately $12.00 per student for the reading and mathematics portions of the test, and about $4.00 per student for the writing samples. These costs did not include the development of test forms and items, nor the statistical analyses necessary to determine test reliability. Second, the fail rate for this district is quite high in comparison to other districts in the state. Statewide, the competency fail rate in grade 10 is from 30-50%* on tests, which though considerably easier, have a passing standard of 65-85% correct. The fail rate for Palos Verdes students at grade eight varies from 55-65% at a passing standard of 50% correct (see Tables 6 and 7). This fail rate can be attributed to several

* Proficiency Assessment in California: 1980 Status Report on Implementation of California Pupil Proficiency Law. (OPER, California State Department of Education, 1980).

TABLE 6.

Percentage of Students Failing One or More Competency Tests
(May, 1981)

Grade*	Percent Failing One or More of 16 Tests
12	19.7 (n=301)
11	28.1 (n=385)
10	61.8 (n=720)
9	55.1 (n=490)

*1981-82 School Year

differences between the Palos Verdes competency program and the majority of programs in other districts in the state:

1. Palos Verdes uses a norm-referenced test (inappropriately) to assess competence; 81% of the districts statewide use teacher-developed tests.
2. Palos Verdes requires students to pass each of 16 subtests to graduate; 80% of the districts require passing scores in reading, writing, and mathematics.
3. The test length for each of the 16 competencies is varied resulting in variable reliability. Table 7 shows the test length and fail rate for the Class of 1985 for each of the 16 tests. It is unclear how these rates would be affected if the tests were equally reliable.
4. District tests do not have clear directions to students, nor are students told the purpose of the test in the written directions.
5. The passing score may not distinguish between those who know the competency and those who do not. It is more likely that a competent student can answer more than 50% of the items and a failing student 50% or less. The actual "score" might better be thought of as a band of scores that reliability distinguishes between those who know the skill and those who do not.

Concerns of the Office of Program Evaluation and Research. Evaluation of the competency testing program will focus on alternatives to the present system, test length, determination of reliability around the passing score for individual forms, and remediation opportunities for students. Specifically, this office is gathering information for cost comparisons among the following:

Table 7.

The Fail Rate for Each Competency Test for the Class of 1985

	Test name	Number of Items	% 8th Grade Failing (1092)
READING	1. Reading comprehension	21	9.2
	2. Vocabulary (8th grade)	15	.8
	3. Reference materials	43	5.0
	4. Interpreting Maps	20	17.5
WRITING	5. 7 sentence paragraph	one paragraph	12.5
	6. Language mechanics	40	1.2
	7. Business and personal letter	one each	8.4
	8. Note taking	one topic given	15. 3
	9. Completing an outline	one passage given	5.4
	10. Sentence combining	3 exercises	20.2
	11. Completing a form	one form	1.2
MATHEMATICS	12. Math computation	31	8.9
	13. Problem solving	37	17.6
	14. Interpret graphs & charts	8	23
	15. Rounding and Estimation	9	6.3
	16. Geometric concepts	5	19.1
	PERCENT FAILING ONE OR MORE TESTS		55.1 (490)

1. Development of an in-house test
2. Purchase of a shelf competency test
3. Commission of Riverside Publications to produce a customized competency test
4. Adaptation of the competency system developed by another district

This office is not as concerned about the writing portion of the competency tests. The major revision of the writing standards consists of updating topics and scoring guidelines to conform to the recent changes in the project.

Cost: How Can Testing and Data Processing Costs be Reduced?

The costs associated with the testing/evaluation program fall into three categories:

1. The cost of test purchase/development
2. Data processing and reporting costs
3. Data storage costs

Test Development Costs. It is not always clear whether it is "cheaper" to purchase or develop tests. Tests developed by the district have the advantage of content validity and match to instructional objectives; the costs of such tests, however, are always higher than professionally-developed measures. These costs are often hidden: they consist of the opportunity costs of paying teachers and administrators to develop tests instead of performing other tasks. Another hidden cost is the amount of time needed to develop a test "in-house." Finally, districts which develop their own tests do not realize the economies of scale made possible by the production of tests for a mass market.

The cost of in-house test development buys much more than a test, however.

Teachers support the test because it reflects their concerns. "Coaching" or teaching to the test items is minimized because teachers believe the test is fair and know how to target their instruction to the test objectives. Test results are valid for making program improvement decisions because the tests match program content and objectives. Participants in the test development and scoring process gain a clearer understanding of the instructional objectives measured by in-house tests because they are forced to analyze them thoroughly in order to produce the test.

Purchase of published tests is less costly than in-house development. Unfortunately, these economies are offset by the lack of district "ownership" in the test and the lack of test validity for making accurate diagnostic decisions about pupils or programs. However, the savings realized by purchasing published tests may be used to assess pupils and programs in less formal but more valid ways (e.g., systematic observation, self-reports, work samples, case studies). Also, teacher ownership of the test may be increased if teachers participate in the test selection process; they may be willing to give up some content validity in exchange for less involvement in the testing program. Table 8 presents the costs of testing under the present system with those of purchasing tests and reports from a publisher.

Data Processing Costs. Data processing and reporting costs are of two sorts: one time only, or developmental costs, and recurring costs. These costs are described in Table 9. Reduction of district data processing costs is directly dependent on the development of an efficient test/evaluation system. This system would reflect the long-term data processing and storage needs of the testing/evaluation program in such a way that programming, processing, and file maintenance costs could be minimized. The elements of a testing/evaluation system require specification of the following:

TABLE 8

Comparative Costs of District-Developed and Published Tests

STANDARDIZED TESTS

District-Developed	Riverside Publishing Company
$5.13 per pupil	$1.30-3.77 per pupil

COMPETENCY TESTS

Present System (Reading and Math)	Iowa Customized(Reading and Math)
$12.73	$3.67 per test=$7.34 per pupil
CTBS Jr. H.S. Tests	CTBS Sr. H. S. Tests
$1.53 per test= $3.06 per pupil	$2.13 per test=$4.26 per pupil

TABLE 9

Costs Associated with Data Processing of Test/Evaluation Information

ONE-TIME (DEVELOPMENTAL) COSTS

Systems analysis and design (including documentation)

Programming (including documentation)

Consultants

Computer time for the testing of the system

RECURRING COSTS

Data preparation (preslugging answer sheets, clerical review of op-scan sheets, etc.)

Personnel: data entry clerk, programmer

Machine time

Supplies: forms, tapes and disk packs, carbon ribbons, etc.

Maintenance programming (see Personnel)

Training of staff

Space for terminal

Transportation

Communication (telephone lines)

Kinds of input:	data, scores, information to be processed
Processing requirements:	data update and editing, report formats, and statistical analyses to be performed
Kinds of output:	number and kinds of reports produced and schedule
Data storage:	organization of student test files, ways to retrieve information from the files, frequency of use, and storage medium
Controls for system:	requirements for obtaining accurate reports, anticipated expansion of system
Feedback for system:	provision for adjustments and revisions of system

The present system used by this district is inadequate on several counts. Programs have not been written with enough flexibility to incorporate reasonable changes in the competency testing program. Statistics produced by the Iowa testing system are improperly calculated and norms tables need updating. There is no discernible file organization (the district had 551 separate disk files for the period April-August, 1981 alone) which has resulted in exorbitant data storage costs ($1.00 per disk track per day; the 551 files took up 1830 disk tracks). File organization (or lack of it) precludes using the computer to "pull out" subsets of students for statistical analysis purposes after initial data processing has been done. Finally, documentation of the system is inadequate. The district possesses no system flowchart, incomplete program flow charts, no programming specifications or copies of the programs, and worse yet, no documentation of what is in the 72 magnetic tapes (each containing several hundred files) kept at RAND. Thus, systems design and documentation work is required whether the district remains with Wayman-RAND, converts completely to the new Burroughs computer, or calls for data processing bids from a test publisher/ outside data processing firm.

Data Storage Costs. The final cost associated with the testing/evaluation system is data storage. A decision as to whether test information should be

available throughout the year, or periodically, and as to how long data are stored is necessary in order to reduce data storage costs. The district spent an average of $800 a month for data storage last year because much of it was kept in an immediately accessable disk file. Periodic use of test information would allow storage on cheaper, magnetic tape.

There are several data processing alternatives for the testing system. Once the system is designed, cost comparisons will be made among the following:

1) Continuation with Wayman-RAND
2) Communication directly with RAND via terminal
3) Purchase of a mini-computer for testing
4) Time sharing
5) Complete conversion to the Burroughs, including the purchase of statistical software
6) Conversion of scoring and reporting to Burroughs, statistical analysis at RAND, USC, or UCLA
7) Contract entire system out to lowest bidder data processing firm
8) Purchase all test statistics, maintain longitudinal files for evaluation purposes on our own computer, statistical analysis on Burroughs or RAND
9) Purchase of test scores, data processing for program evaluation done manually
10) Data storage and processing at RAND, with rental of a terminal from RAND for data processing as needed (it can be brought to the district).

Cost reduction is not only desirable, but absolutely essential. Whether reduction occurs as increased efficiency or decreased cost depends on the information requirements of the district. Table 10 presents potential long-term

data collection and processing needs. Changes in federal and state law may reduce some of these needs. Changes in district programs and priorities may expand them. This list, nevertheless, can be used to begin work on the systems design so that cost-benefit analyses may be made.

TABLE IO

LONG-TERM INFORMATION NEEDS OF DISTRICT

PROGRAM		TIME NEEDED	STORAGE REQUIREMENTS
Title I Bilingual Special Education	Pre-post test norm-referenced scores for reading, language arts, mathematics	Feb.-March	3-12 years
Competency Test	Pass/fail record for each student on each test Percentage of students passing each test at grades 5, 8, 9-12	Grades 5,8:Spring Grade 12: Sept. Jan. Mar. May	5-7 years
Writing Standards Test (including competency portions)	Number correct score on writing piece Pass/fail score at grades 5, 8, 11	Spring	0-12 years
Reading Specialist Resource Teacher Special Day Class GATE Modified	Separate files for each of these programs containing student names and test information (Iowa, competency, writing standards, other)	Oct.Feb.June	3-12 years
Mainstreamed students: FEB, Special Ed., Reading, Modified Title I	Code to identify so success in regular program can be studied via district tests, grades, other	Unknown	3-12 years
California Assessment Program	School and district percent correct scores District and school percentile ranks	Fall	3-6 years
Any New Programs (Math,Revised Modified, Writing,Social Studies, Science, etc.)	Pre-posttest information for a sample of students in each class and school	Unknown	1-4 years